

Association of American Publishers  
Meeting Wednesday, July 28, 1999

Patricia Schroeder  
President & CEO, AAP

Allan R. Adler  
VP, Legal & Government Affairs, AAP

Joel Baron  
Publisher, NEJM

Bob Bovenschulte  
Director, Publications, ACS

Pieter Bolman  
President, Academic Press

Brian Crawford  
VP & Gen Mgr, Life & Med Sci, John Wiley

Charles Ellis  
Senior Advisor, John Wiley

David Lipman  
Judith Turner

Joel Baron opened by saying that the group was "very happy that NIH has taken this initiative" and that they "understand how important it is to cooperate." He said that now that publishers had gotten to the point of putting their publications on line, they were asking what they do with them. "What is the power in this setup."

He said they were looking for "clarification and cooperation—how might we work together so the outcomes are reasonable for us as well."

David said that what the NIH would do is "incredibly minimal"—create a repository for peer-reviewed and screened material. He said that the NIH

would not be involved in choosing the material, although it might be involved in choosing the groups that could participate. He said that although it has not been finally decided, NIH is leaning toward having other groups make the decisions about screening, too. "The role of the NIH is passive," he said. He said the NIH would develop tools to take a DTD and create templates to format material, to facilitate the front end. The repository would allow searching, reading, linking, and archiving. He defined three important aspects of archiving:

1. Safety. Mirror sites would be updated immediately. The advisory board would decide where the mirror sites would be.
2. Proper documentation for each publication's DTD and a program to convert it to NIH's DTD so it can be moved along as the technology advances.
3. Physical storage. He said the NIH's other NLM databases were copied and stored under a mountain in Pennsylvania, and this would be, too. He pointed out that archiving is important because "for people in life science, content that is 30 years old has value." It is all part of the active system, he said.

He was asked whether the NIH would work with groups like JSTOR, and he said he assumed so.

He was asked whether the NIH would do retrospective conversion, and said that Dr.V. hoped that money would come from philanthropies, drug companies, and high-tech companies. He said that some companies had already contacted him expressing interest.

In answer to a question, he said it would be international. He said EMBO and EMBL had already decided they must be involved. He said he wanted to "get some momentum—make sure there is some 'there' there—before it gets too encumbered" with an international bureaucracy. Several of the AAP reps nodded, as they have been involved with the international DOI effort that has bogged down in bureaucracy. David said the NIH would be a leader because it started the program, but if there is not international involvement it would hurt the program.

He said that the idea that the NIH would do peer review has been dropped. "The government shouldn't get into this," he said. He said that EMBO's idea

is that two scientists do screening, and do an up or down vote without making suggestions for change. EMBO, he said, had tough criteria; his idea was that screening eliminates porn, slander, advertising and crackpots. "But we will not make those decisions; I feel comfortable with that." He said that the other advantage is that "no one knows what a life-science preprint server should look like," and this would allow experimentation.

There was discussion about "unexciting but useful science," and David said that journals would have to decide whether to have a "second-level" publication, or maybe a group of journals could do it together to avoid branding. He also discussed the possibility that a single review board could make a decision for a set of journals; someone questioned the effect of that on diversity.

David said the goal is a community knowledge base with no barriers to access, and explained the two ways he saw journals creating economic models:

- Backward thinking, in which a journal looks at current revenue streams and usage patterns, and decides what is endangered, and then assumes it must make up that money. David said the NIH's role would be to allow authors to pay a fee out of their grant funds to put their work online, and he said they are talking to other agencies to get them to allow it.
- Forward thinking, in which a journal looks at its expenses in producing a journal, looks at who buys it, and questions what happens if it is online only. He said that the costs could be about \$400/article.

Joel Baron challenged that, saying that sometimes they had to pay for peer review if it was outside their membership, and that it cost an average of \$10,000/published article for peer review, taking into account that they still have to review the 90% of articles they don't accept.

David said that some journals will always be in paper, but ultimately the author would make the decision of where to publish—perhaps in part based on the cost of putting the work online. He acknowledged that the top journals might charge more because they do more. (He said the plant physiologists are considering a system in which reviewers can determine if the author should be charged extra for editing.)

He pointed out that all the journals are saying that it is costing more money to be online, but they are not making more money. (Heads nodded around the table.)

David said that the NIH economic goal is "if we are going to spend billions funding research, we want more people to see it."

Charles Ellis clarified: "The arena of free-market choices is moving from the user to the author" with the costs reflected in the author's charges. He questioned how the third-world scientist would fare. David pointed out that some journals now make dispensations on page charges, and that the fraction of articles from the third world is very small. He also said that while the European funding agencies won't pay page charges, they may pay an "access" charge [to allow access to the paper].

Brian Crawford asked whether authors could choose, article by article, to pay the cost of being on E-biomed. David said that currently the only parameter we are considering is delay, but "in the spirit of experimentation I am open to that." He said that unless the fee was "tremendous," most authors would want to pay. "For the author, it is unbelievable that everyone will get access; nobody will write a paper and not cite your work because they can't get access to it."

In answer to a question, David said that it had to be a central database because "to make a corpus of material useful, you need to be able to make changes, that that could not be done if it is scattered across many sites." He said several publishers have already agreed to that, and he intends to build it and show that it works.

Bob Bovenschulte asked about a national site license. "That puts NIH in the position of saying 'this we will pay for and this we will not,'" David said. (There was nodding around the table.) "The advantage of this is that even if it is a stunning success, it will be slow." He said the pace will give companies time to plan for the transition.

Asked about the costs of the infrastructure, David said it would be \$2-million to \$3-million, and that NIH was hoping that other agencies would put money into it. "In the end, the amount of the investment NIH will do will be

minuscule," he said. The most important thing the NIH was bringing to the effort was the bully pulpit. "We got people thinking about it."

He was asked who would be allowed to participate. "This is a critical issue, and we want your feedback," he said. "Any journal already indexed in the major A&I's, and any new group that has on its editorial board three or more members who are PI's with major funders." He said the point was to minimize "total trash" and yet be inclusive. If there were groups that were problematic, their participation would be decided by an outside group—the Advisory Board or the National Academy. He added that any journal in E-biomed would be indexed in Pub Med, but maybe not in MESH or the Index Medicus, because they are done by hand.

Joel Baron asked whether in E-biomed peer review becomes a commodity—a problem for journals that reviewed many but accepted few. David said that submission and acceptance fees could be separated.

David asked for the group's help with copyright. "What should E-biomed's rights be?" he asked. "We want to work with you on that." Pat Schroeder said that Allan Adler on her staff was a copyright lawyer, and would be available. David raised issues not only of fair use but of use of datasets, the international component of copyright, and other issues.

David said that NIH is working on a programming language that writes HTML on the fly, with C and C++ scripts, based on SGML (where we know the DTD). It gets at the storage-vs-presentation issues, and it will be made available to the publishing community. The advantage is that when E-biomed is changed/upgraded, you don't have to reconvert the HTML.

It was suggested that NIH put up on its Web site the technological problems it was trying to solve, and what stage we are at, so that people could alert us if it's been solved elsewhere. For instance, Joel Baron said, there is an Elsevier/Blackwell/Springer DTD standard already.

In answer to a question, David said that the E-biomed repository would, by January, search Pub Med, and link to the article at the publisher's and E-biomed sites. It would include links from the article to the cites, and from the cites back to the article. There will soon be a way to pull out useful

terms, gene symbols, etc. and go to the primary database from them. There will be a program that will automatically recognize highly useful terms. It will also be possible to search parts of the paper, for instance looking at the "methods" section and asking for other papers that use similar methods.

Joel Baron asked what would happen if Congress were to decide that the NIH can't publish stem-cell research. "I have more confidence in Congress than that," David said. "It is one thing to say don't do stem-cell research, and another thing to say 'burn the books.'"

The group suggested that NIH "recast" the original proposal to explain it better, perhaps identifying elements that are fixed, and where there is still room for discussion. David said he might take a crack at that. They said that the weakest part of the proposal was the business model. David answered that he did not think the NIH should propose a business model. "It's not our territory," he said.